WELTORGANISATION FUR GEISTIGES EIGENTUM Internationales Būro

INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation 6:

C12N 15/12, C07K 14/47, 16/18, C12N 15/11, G01N 33/50, A01K 67/00, C12N 1/21, 5/10

(11) Internationale Veröffentlichungsnummer: WO 96/35784

A2

(43) Internationales Veröffentlichungsdatum:

14. November 1996 (14.11.96)

(21) Internationales Aktenzeichen:

PCT/EP96/01818

(22) Internationales Anmeldedatum:

2. Mai 1996 (02.05.96)

(81) Bestimmungsstaaten: CA, JP, MX, US, europäisches Patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

(30) Prioritätsdaten:

*

195 16 776.7

10. Mai 1995 (10.05.95)

DE

Veröffentlicht

Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.

BOEHRINGER INGELHEIM INTERNATIONAL GMBH [DE/DE]; Postfach 200, D-55216 Ingelheim am Rhein (DE).

(71) Anmelder (für alle Bestimmungsstaaten ausser US):

(72) Erfinder: und

(75) Erfinder/Anmelder (nur für US): JENUWEIN, Thomas [DE/AT]; Barichgasse 21/27, A-1030 Wien (AT). LAIBLE, Götz [DE/AT]; Costagasse 9/13, A-1150 Wien (AT).

(74) Gemeinsamer Vertreter: BOEHRINGER INGELHEIM IN-TERNATIONAL GMBH; Postfach 200, D-55216 Ingelheim am Rhein (DE).

(54) Title: CHROMATIN-REGULATOR GENES

(54) Bezeichnung: CHROMATIN-REGULATORGENE

(57) Abstract

The invention concerns the deregulation of chromatin-regulator genes which have an SET domain, such deregulation being of importance in certain cancer conditions. These genes, in particular the SET domains as such, can be used in the diagnosis and therapy of such conditions.

(57) Zusammenfassung

Die Deregulation von Chromatinregulatorgenen, die eine SET-Domäne aufweisen, spielt eine Rolle bei bestimmten Krebserkrankungen. Diese Gene können somit, insbesondere die SET-Domäne als solche, als Grundlage für die Diagnose und Therapie dieser Erkrankungen dienen.

4

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AM	Armenien	GB	Vereinigtes Königreich	MX	Mexiko
AT	Österreich	GE	Georgien	NE	Niger
ΑÜ	Australien	GN	Guinea	NL	Niederlande
BB	Barbados	GR	Griechenland	NO	Norwegen
BB	Belgien	HU	Ungarn	NZ	Neuseeland
BF	Burkina Faso	IB	Irland	PL	Polen
BG	Bulgarien	IT	Italien	PT	Portugal
BJ	Benin	JP	Japan	RO	Ruminien
BR	Brasilien	KR	Kenya	RU	Russische Föderation
BY	Belarus	KG	Kirgisistan	SD	Sudan
CA	Kanada	KP	Demokratische Volksrepublik Korea	SE	Schweden
CF	Zentrale Afrikanische Republik	KR	Republik Korea	SG	Singapur
CG	Kongo	KZ	Kasachstan	81	Slowenien
CH	Schweiz	LI	Liechtenstein	SK	Slowakei
CI	Côte d'Ivoire	LK	Sri Lanka	SN	Senegal
CM	Kamerun	LR	Liberia	8Z	Swasiland
CN	China	LK	Litmen	TD	Techad
CS	Tachechoslowakei	LU	Luxemburg	TG	Togo
CZ	Tschechische Republik	LV	Lettland	TJ	Tadachikistan
DE	Deutschland	. MC	Мовасо	TT	Trinidad und Tobago
DK	Dänemark	MD	Republik Moldau	UA	Ukraine
KB	Estland	MG	Madagaskar	UG	Uganda
E8	Spanien	ML.	Mali	US	Vereinigte Staaten von Amerika
FI	Finnland	MN	Mongolei	UZ	Usbekistan
FR	Prankreich	MR	Mauretanien	VN	Vietnam
GA	Gabon	MW	Malawi		

Chromatin-Regulatorgene

Die vorliegende Erfindung bezieht sich auf Gene, die eine Rolle bei der strukturellen und funktionellen Regulation von Chromatin spielen, und ihre Verwendung für die Therapie und Diagnostik.

Die funktionelle Organisation von eukaryotischen Chromosomen in Centromere, Telomere sowie in eu- und heterochromatische Regionen stellt einen entscheidenden Mechanismus zur Gewährleistung der genauen Replikation und Verteilung der genetischen Information bei jeder Zellteilung dar. Im Gegensatz dazu sind Tumorzellen häufig durch chromosomale Rearrangements, Translokationen und Aneuploidie charakterisiert (Solomon et al., 1991; Pardue, 1991). Obwohl die Mechanismen, die zu einer erhöhten Chromosomeninstabilität in Tumorzellen führen, noch nicht geklärt sind, haben es in jüngster Zeit eine Reihe von experimentellen Systemen, beginnend mit telomerischen Positionseffekten in Hefe (Renauld et al., 1993; Buck und Shore, 1995; Allshire et al., 1994), über Positionseffekt-Variegation (PEV) in Drosophila (Reuter und Spierer, 1992), bis zur Analyse von Translokationsbruchpunkten in humanen Leukämien (Solomon et al., 1991; Cleary, 1991), ermöglicht, einige chromosomale Proteine zu identifizieren, die an der deregulierten Proliferation ursächlich beteiligt sind.

Erstens wurde festgestellt, daß die Überexpression einer verkürzten Version des SIR4-Proteins zu einer verlängerten Lebensdauer in Hefe führt (Kennedy et al., 1995). Da SIR-Proteine zum Entstehen multimerer Komplexe an den stillen "Mating Type Loci" und am Telomer beitragen, könnte es sein, daß überexprimiertes SIR4 mit diesen heterochromatinartigen Komplexen interferiert, was schließlich zu einer unkontrollierten Proliferation führt. Diese Annahme steht im Einklang mit der Häufigkeit des Auftretens einer deregulierten Telomerenlänge in den meisten humanen Krebsarten (Counter et al., 1992).

Zweitens wurden mittels genetischer Analysen von PEV in Drosophila eine Reihe von Genprodukten identifiziert, die die Chromatinstruktur an heterochromatischen Positionen und innerhalb des homeotischen Genclusters verändern (Reuter und Spierer, 1992). Mutationen einiger dieser Gene, wie modulo (Garzino et al., 1992) und polyhomeotic (Smouse und Perrimon, 1990), können in Drosophila die deregulierte Zellproliferation oder den Zelltod verursachen. Drittens wurden Säugetierhomologe von sowohl Aktivatoren (trithorax oder trx-Gruppe) als auch Repressoren (z.B. polycomb oder Pc-Gruppe) der Chromatinstruktur von homeotischen Drosophila-Selektorgenen beschrieben. Unter diesen hat sich vom humanen HRX/ALL-1 (trx-Gruppe) gezeigt, daß es an der durch Translokation induzierten Leukamogenese beteiligt ist (Tkachuk et al., 1992; Gu et al., 1992), und daß die Überexpression des Maus-bmi (Pc-Gruppe) zum Entstehen von Lymphomen führt (Haupt et al., 1991; Brunk et al., 1991; Alkema et al., 1995). Ein Modell für die Funktion chromosomaler Proteine läßt darauf schließen, daß diese multimere Komplexe bilden, welche in Abhängigkeit von der Ausgewogenheit zwischen Aktivatoren und Repressoren im Komplex den Kondensationsgrad der umliegenden Chromatinregion bestimmen (Locke et al., 1988). Eine Verschiebung dieses Gleichgewichts durch Überexpression einer der Komponenten des Komplexes zeigte eine Neuverteilung von eu- und heterochromatischen Regionen (Buck und Shore,

1995; Reuter und Spierer, 1992; Eissenberg et al., 1992). Dieser Dosiseffekt kann die Chromatinstruktur an vorbestimmten Loci destabilisieren, was letztlich zu einem Übergang vom normalen zum transformierten Zustand führt.

Trotz der Charakterisierung von HRX/ALL-1 und bmi als Protoonkogene, die die Chromatinstruktur verändern können, ist das Wissen über Säugetiergenprodukte, welche mit Chromatin wechselwirken, noch sehr beschränkt. Im Gegensatz dazu wurden durch genetische Analysen von PEV in Drosophila ca. 120 Allele für Chromatinregulatoren beschrieben (Reuter und Spierer, 1992). Kürzlich wurde eine carboxyterminale Region mit Ähnlichkeit in der Sequenz identifiziert, die einem positiven (trx, trx-Gruppe) und einem negativen (E(z), Pc-Gruppe) Drosophila-Chromatinregulator (Jones und Gelbart, 1993) gemeinsam ist. Darüberhinaus ist dieser Carboxyterminus auch in Su(var)3-9, einem dominanten Supressor der Chromatinverteilung in Drosophila, konserviert (Tschiersch et al., 1994).

Bei der vorliegenden Erfindung wurde von der Überlegung ausgegangen, daß diese als "SET" bezeichnete Proteindomäne (Tschiersch et al., 1994) wegen ihrer evolutionären Konservierung und dem Vorhandensein in antagonistischen Genprodukten eine neue Genfamilie entwicklungsgeschichtlich wichtiger Säugetier-Chromatinregulatoren definiert. Darüberhinaus trägt die Charakterisierung weiterer Mitglieder der Gruppe von SET-Domäne-Genen, neben HRX/ALL-1, zur Aufklärung der Mechanismen bei, die dafür verantwortlich sind, daß strukturelle Veränderungen im Chromatin zur malignen Transformation führen können.

Der vorliegenden Erfindung lag somit die Aufgabe zugrunde, humane Chromatinregulatorgene zu identifizieren, ihre Funktion aufzuklären und sie für die Diagnostik und Therapie einzusetzen.

Zur Lösung dieser Aufgabe wurde zunächst die Sequenzinformation der SET-Domäne benutzt, um aus humanen cDNA-Banken die zu den SET-Domäne-Genen von Drosophila homologen humanen cDNAs zu erhalten. Es wurden zwei cDNAs erhalten, die Humanhomologe von E(z) bzw. Su(var)3-9 darstellen; die entsprechenden humanen Gene wurden als EZH2 und SUV39H bezeichnet (vgl. unten); außerdem wurde eine variante Form von EZH2 identifiziert, die als EZH1 bezeichnet wurde.

Die vorliegende Erfindung betrifft somit DNA-Moleküle, enthaltend eine für ein Chromatinregulator-Protein, das eine SET-Domäne aufweist, kodierende Sequenz oder eine Teilsequenz davon, dadurch gekennzeichnet, daß sie die in Fig. 6 oder in Fig. 7 dargestellte Nukleotidsequenz aufweisen. Die erfindungsgemäßen DNA-Moleküle werden im folgenden auch als "erfindungsgemäße Gene" bezeichnet.

Die erfindungsgemäßen Gene tragen die Bezeichnung EZH2 und SUV39H, sie wurden ursprünglich als "HEZ-2" und "H3-9" bezeichnet.

Die Erfindung betrifft in einem weiteren Aspekt die von diesen Genen abgeleiteten cDNAs, einschließlich ihrer degenerierten Varianten; Mutanten, die für funktionelle Chromatinregulatoren kodieren sowie Varianten, die auf eine Genduplikation zurückzuführen sind; ein Beispiel dafür ist EZH1, dessen Teilsequenz im Vergleich mit EZH2 in Fig. 8 dargestellt ist.

Zur Lösung der im Rahmen der vorliegenden Erfindung gestellten Aufgabe wurde im einzelnen wie folgt vorgegangen: Ausgehend von der Sequenzinformation von der konservierten SET-Domäne, wurde unter reduzierter Stringenz eine humane B-Zell-spezifische cDNA-Bibliothek mit einer gemischten Drosophila-DNA-Sonde, die für die SET-Domänen von E(z) und Su(var)3-9kodiert, gescreent. Aus 500.000 Plagues wurden 40 primäre Phagen ausgewählt. Nach zwei weiteren Runden Screening zeigte sich, daß 31 Phagen für authentische E(z)-Sequenzen kodieren und daß fünf Phagen E(z)-Varianten darstellen. Im Gegensatz dazu hybridisierten nur zwei Phagen mit der Sonde, die die SET-Domäne von Su(var)3-9 allein enthielt. Die Phageninserts wurden mittels Polymerasekettenreaktion (PCR) amplifiziert und mittels Restriktionskartierung und Teilsequenzierung analysiert. Repräsentative cDNA-Inserts wurden subkloniert und über ihre ganze Länge sequenziert. Die 5'-Enden wurden isoliert, indem positive Phagen noch einmal mit 5'-DNA-Sonden gescreent wurden, worauf nach Subklonierung komplette cDNAs erhalten wurden.

Die komplette, für das humane Homologe von E(z) kodierende cDNA wurde als EZH2 und die DNA, die für das humane Homologe von Su(var)3-9 kodiert, wurde als SUV39H bezeichnet. Insgesamt beträgt die Identität der Aminosäuren zwischen Drosophila und den humanen Proteinen 61 % für EZH2 und 43 % für SUV39H, wobei die C-terminale SET-Domäne sehr hoch konserviert ist (88 % für EZH2 und 53 % für SUV39H). Der Sequenzvergleich zeigte weitere deutliche Homologieregionen, z.B. eine Cystein-reiche Domäne in EZH2 und eine Chromo-Box in SUV39H. (In POLYCOMD wurde gezeigt, daß die Chromo-Box die für die Wechselwirkung zwischen DNA und Chromatin wesentliche Domäne ist; Messmer et al., 1992). Im Gegensatz dazu fehlen die

207 Aminosäuren, die das aminoterminale GTP-Bindungsmotiv des *Drosophila*-Proteins enthalten, im humanen Homologen *SUV39H*.

Der Vergleich der Aminosäure-Sequenzen zwischen den Drosophila- und den humanen Genen ist in den Fig. 1 und 2 dargestellt:

Fig. 1 zeigt den Vergleich der Aminosäure-Sequenz zwischen EZH2 und von Drosophila Enhancer von zeste (E(z)). Identische Aminosäuren der konservierten carboxyterminalen SET-Domäne (schattierte Box) und der Cys-reichen Region (Cys-Reste sind hervorgehoben) sind gezeigt. Die mutmaßlichen Kernlokalisierungssignale sind unterstrichen.

Fig. 2 zeigt den Vergleich der Aminosäure-Sequenz zwischen dem humanen Homologen SUV39H und Drosophila Su(var)3-9. Identische Aminosäuren der konservierten carboxyterminalen SET-Domäne (schattierte Box) und der Chromo-Domäne (dunkler schattierte Box) sind dargestellt. Die mutmaßlichen Kernlokalisierungssignale sind unterstrichen. Oben in der Figur ist eine schematische Übersicht der beiden Proteinstrukturen dargestellt, die zeigt, daß im humanen Homologen 207 Aminosäuren am N-Terminus fehlen.

Da translationale Consensussequenzen in Umgebung des Start-ATG der humanen SUV39H-cDNA auch an der entsprechenden internen Position in Su(var)3-9 vorhanden sind, dürfte das Drosophila-Protein zusätzliche Exons enthalten, die in einem späteren Stadium der Evolution für die Funktion entbehrlich wurden. (Die Richtigkeit dieser Hypothese kann bestätigt werden, indem die humane SUV39H-cDNA und komplette oder am 5'-Ende verkürzte cDNAs von

Su(var)3-9 in Drosophila exprimiert werden.

Darüberhinaus wurde eine weitere cDNA der Bezeichnung

MG-44 beschrieben

(s. unten), der ebenfalls das 5'-Ende des Drosophila-Gens fehlt.) Zusätzlich zur humanen cDNA von SUV39H wurde auch der homologe Locus in der Maus (Suv39h; siehe unten) isoliert, dessen Sequenzanalyse und Promotorstruktur eindeutig die aminoterminale Verkürzung säugerhomologer Gene im Vergleich zu Drosophila Su(var)3-9 bestätigt.

Im Rahmen der vorliegenden Erfindung durchgeführte DNA-Blot-Analysen deuten darauf hin, daß säugerhomologe Gene von Su(var)3-9 in der Maus und im Menschen von einzelnen Loci repräsentiert werden, während säugerhomologe Gene von E(z) in der Maus und im Menschen durch zwei getrennte Loci kodiert sind. Der zweite humane Locus (als EZH1 bezeichnet) wurde auch durch Charakterisierung einer kleinen Anzahl von cDNA-Varianten, die sich in ihren 3'-flankierenden Sequenzen von der Mehrzahl der aus der humanen cDNA-Bibliothek isolierten Klone unterscheiden, bestätigt. Die Unterschiede zwischen EZH2 und EZH1 im sequenzierten Bereich sind in Fig. 8 dargestellt: Die SET-Domäne von EZH1 zeigt gegenüber EZH2 Mutationen; außerdem trägt die von uns isolierte EZH1-Variante (sehr wahrscheinlich eine aberrant gespleißte cDNA) ein sich im Leserahmen befindliches Stopcodon, welches das Protein um die 47 C-terminalen Aminosäuren verkürzt. In Fig. 8 ist die Nukleotidsequenz der EZH2-cDNA von Position 1844 bis 2330 in der jeweils oberen Zeile dargestellt, wobei die 5' Spleißstelle und das potentielle Stopcodon unterstrichen sind. Um eine Teilsequenz der cDNA der EZH1-Variante der EZH2-Sequenz zuzuordnen, wurde das gap-Programm des Wisconsin GCG Netzwerkservice verwendet. Das vorzeitige Stopcodon in

EZH1 (Position 353) ist unterstrichen. Sequenzen, die für die konservierte SET-Domäne kodieren, sind hervorgehoben. Außerdem ist das 3'-Ende (Position 151 in EZH1) des aberranten Transkripts B52 (s. unten) dargestellt. Über die verfügbare Sequenz zeigte sich B52 zu 97 % identisch mit EZH1 und zu 72 % identisch mit EZH2. Der Sequenzvergleich von EZH1 mit EZH2 sowie die Feststellung, daß beim Menschen und der Maus zwei getrennte E(z)-homologe Loci auftreten, lassen darauf schließen, daß bei Säugern eine Genduplikation aufgetreten ist.

In einem Vergleich mit cDNA-Sequenzen in der Datenbank GeneBank wurde überraschend festgestellt, daß bestimmte in der Datenbank eingetragene cDNA-Teilsequenzen, die von aberranten Transkripten in Tumorgeweben abgeleitet sind, mutierte Versionen der erfindungsgemäßen cDNAs darstellen:

Einerseits war auf der Suche nach BRCA1, einem Gen, das für Brust- und Eileiterkrebs prädisponiert, eine partielle cDNA-Sequenz mit 271 Nukleotiden der Bezeichnung B52, die für eine mutierte Variante der SET-Domäne kodiert, isoliert und auf dem humanen Chromosom 17q21 kartiert worden (Friedman et al., 1994). Es wurde im Rahmen der vorliegenden Erfindung überraschend festgestellt, daß B52 97 % Identität mit der erfindungsgemäßen BZH1-cDNA-Variante aufweist (vgl. oben); möglicherweise stellt EZH1 ein Gen dar, dessen Reaktivierung bei der deregulierten Proliferation eine Rolle spielt.

Andererseits war eine cDNA (2.800 Nukleotide; MG-44) vom humanen Chromosom Xpl1 isoliert worden (Geraghty et al., 1993), einer Region, die für degenerative Störungen der Netzhaut und Synovialsarkome

prädisponiert. Es wurde überraschend festgestellt, daß diese cDNA 98 % Identität mit der erfindungsgemäßen SUV39H-cDNA aufweist.

Die im Rahmen der vorliegenden Erfindung bereitgestellten neuen Gene ermöglichen es somit, einen Zusammenhang zwischen bestimmten Krebserkrankungen und Mutationen in Chromatinregulatoren herzuleiten; im Falle der MG-44-cDNA konnte, da diese mehrere Punktund Frameshift-Mutationen aufweist, welche die Chromound SET-Domänen unterbrechen, erst anhand der erfindungsgemäßen SUV39H-cDNA ein Zusammenhang zwischen Su(var)3-9 und MG-44 aufgeklärt werden.

Neben den bereits genannten Sequenzen sind in der Sequenzdatenbank GeneBank als weitere humane Mitglieder der SET-Proteinfamilie das gut dokumentierte humane Homologe von Drosophila trx, HRX/ALL-1 (Tkachuk et al., 1992; Gu et al., 1992) eingetragen, ferner ein Gen unbekannter Funktion der Bezeichnung G9a, das im humanen Major Histocompatibility Complex (Milner und Campbell, 1993) vorhanden ist, drittens eine nicht publizierte cDNA (KG-1), die aus unreifen myeloiden Tumorzellen isoliert wurde (Nomura et al., 1994). Während G9a derzeit das einzige humane Gen mit einer SET-Domane ist, für das bisher keine mutierte Version bekannt ist, trägt KG-1 eine Insertion von 342 Aminosauren, welche die SET-Domane in eine Amino- und eine Carboxyterminale Hälfte spaltet. Wahrscheinlich stellt diese KG-1-cDNA eine aberrant gespleißte Variante dar, da sich 5' und 3' Konsensus-Spleißstellen and beiden Enden der Insertion finden. Insgesamt sind vier der fünf derzeit bekannten humanen Mitglieder der SET-Proteinfamilie Änderungen unterworfen, die alle die SET-Domane mutieren (HRX/ALL-1, EZH1/B52, SUV39H/MG-44

und KG-1). Darüberhinaus wurden in drei Fällen die entsprechenden humanen Genloci in der Nähe von Translokationsbruchpunkten oder instabilen chromosomalen Regionen kartiert (HRX/ALL-1, EZH1/B52 und SUV39H/MG-44). In Fig. 3 sind die aberranten Transkripte von humanen SET-Domäne-Genen dargestellt. Links in der Figur ist die Lage der fünf derzeit bekannten SET-Domäne-Gene auf dem jeweiligen Chromosom angegeben. U.a. sind die drei Gene (HRX/ALL-1, EZH1/B52 und SUV39H/MG-44), für die aberrante cDNAs auf Translokationsbruchpunkten oder instabilen Chromatinregionen kartiert wurden, dargestellt. Vier der fünf dargestellten SET-Domänen Gene weisen Mutationen auf, die alle die Carboxyterminale SET-Domane unterbrechen, welche in der Figur durch die dunkle Box dargestellt ist. Eine Translokation verbindet die aminoterminale Hälfte von HRX mit einer nicht korrelierten Gensequenz, die als gepunktete Box mit der Bezeichnung ENL dargestellt ist. Mutationen und ein vorzeitiges Stopcodon verändern die SET-Domäne von EZH1/B52. Punkt- und Frameshift-Mutationen unterbrechen die Chromo- und SET-Domane in MG-44. Eine große Insertion spaltet die SET-Domane von KG-1 in zwei Hälften. Aberrante Transkripte sind derzeit für G9a nicht bekannt. Das Cystein-reiche Cluster in B52 ist als gepunktete Box dargestellt; in HRX/ALL-1 sind die Homologieregion zu Methyltransferasen als schraffierte Box und die A/T-Haken als vertikale Linien dargestellt. Die Namen der jeweiligen authentischen Gene sind rechts in der Figur angegeben.

Die Tatsache, daß ein Säugetiergen der SET-Proteinfamilie, HRX/ALL-1, mit translokationsinduzierter Leukämogenese im Zusammenhang gebracht wurde (Tkachuk et al., 1992; Gu et al., 1992), ist ein starker Hinweis dafür, daß Proteine mit der SET-Domäne 11

nicht nur wichtige Regulatoren der Entwicklung sind, die chromatinabhängige Veränderungen der Genexpression mitbestimmen, sondern daß sie, nach Mutation, auch die normale Zellproliferation stören.

Da alle bisher beschriebenen Mutationen die Primärstruktur der SET-Domäne unterbrechen, liegt die Vermutung nahe, daß es die SET-Domäne als solche ist, die eine entscheidende Rolle im Übergang vom normalen zum transformierten Zustand spielt. Eine wichtige Funktion für die SET-Domäne kann ferner aufgrund ihrer evolutionären Konservierung in Genprodukten, die von der Hefe bis zum Menschen auftreten, vermutet werden.

Fig. 4 zeigt die evolutionäre Konservierung von SET-Domäne-Proteinen: Unter Verwendung des tfasta-Programms des Wisconsin GCG Netzwerkservice wurden Proteine und offene Leserahmen mit Homologie zur SET-Domäne identifiziert. In der Figur ist eine repräsentative Auswahl von Hefe bis zum Menschen gezeigt. Ziffern geben Aminosäuren an. Die Carboxyterminale SET-Domäne ist durch eine schwarze Box dargestellt, Cys-reiche Regionen durch eine dunkel gepunktete Box, das GTP-Bindungsmotiv in Su(var)3-9 durch eine hellgepunktete Box und die Chromo-Domane von Su(var) 3-9 und H3-9 durch eine offene Box mit hellen Punkten. Eine Region mit Homologie zu Methyltransferase (trx und HRX) ist als schraffierte Box dargestellt. A/T-"Haken" ("A/T Hooks") sind durch vertikale Linien dargestellt. Eine weitere Ser-reiche Region (S in C26E6.10) und eine Glu-reiche Region (E in G9a) oder Ankyrin-Repeats (ANK in G9a) sind ebenfalls hervorgehoben. YHR119 (GeneBank Accession No. U00059) und C26E6.10 (GeneBank Accession No. Ul3875) sind offene Leserahmen von kürzlich in die Datenbank eingetragenen Cosmiden ohne eine funktionelle Charakterisierung. Prozente geben die Gesamtheit der

Aminosäure-Identitäten zwischen den humanen und den Drosophila-Proteinen an.

Fig. 5 zeigt die Übereinstimmung an Aminosäuren der SET-Domane. Die SET-Domane der in Fig. 4 dargestellten Gene wurde unter Verwendung des Pileup-Programms des Wisconsin GCG Netzwerkservice angeordnet. Für den Vergleich der KG-1-SET-Domäne wurde vor dem Pileup die große Aminosäure-Insertion, die die SET-Domäne in zwei Hälften spaltet, entfernt. Aminosäure-Positionen, die 8 von 10 Übereinstimmungen aufweisen, sind hervorgehoben.

Aufgrund der im Rahmen der vorliegenden Erfindung festgestellten Kriterien erweist sich eine Beteiligung der Gene, die eine SET-Domäne haben, an der chromatinabhängigen Entstehung der deregulierten Proliferation; diese Gene bzw. die davon abgeleiteten cDNAs sowie Teil- und mutierte Sequenzen können somit bei der Therapie und Diagnose von Erkrankungen, die auf eine derartige Proliferation zurückzuführen sind, eingesetzt werden:

Unterschiede im Transkriptionsniveau von SET-Domäne-RNAs zwischen normalen und transformierten Zellen können als diagnostischer Parameter für Erkrankungen herangezogen werden, in denen die Expression von SET-Domāne-Genen dereguliert ist:

So können Oligonukleotide, kodierend für die SET-Domäne als solche bzw. Teilabschnitte davon, als diagnostischer Marker verwendet werden, um bestimmte Krebsarten, in denen die SET-Domäne mutiert ist, zu diagnostizieren. Für die detaillierte Analyse der Funktion der erfindungsgemäßen cDNAs oder Abschnitten davon im Hinblick auf den diagnostischen Einsatz von

SET-Domäne-Gensequenzen wurden im Rahmen der vorliegenden Erfindung die homologen Maus-cDNAs von EZH1 (Ezh1) und SUV39H (Suv39h) isoliert. Bei Verwendung einer Maus-spezifischen, für die SET-Domäne kodierenden, DNA-Sonde in "RNAse protection" Analysen zur Untersuchung der Ezhl-Genaktivität während der normalen Mausentwicklung zeigte sich ein eher breites Expressionsprofil, das ähnlich dem von bmi (Haupt et al., 1991) ist. Die mit den Maussequenzen durchgeführten Analysen werden mit Humansequenzen ausgeweitet, um die RNA-Mengen zwischen unreifen Vorläuferzellen, Tumorzellen und differenzierten Zellen in verschiedenen humanen Zellkultursystemen zu vergleichen. Um festzustellen, ob sich die SET-Domäne dementsprechend als diagnostischer Tumormarker für spezifische Krebserkrankungen oder als genereller diagnostischer Parameter eignet, können gängige Methoden zur Bestimmung der RNA-Konzentration, die in einschlägigen Laborhandbüchern (Sambrook, J., Fritsch, E.F. und Maniatis, T., 1989, Cold Spring Harbor Laboratory Press) beschrieben sind, wie Northern Blot, S1-Nuclease-Protection-Analyse oder RNAse-Protection-Analyse, verwendet werden.

Um die Häufigkeit zu untersuchen, mit der die SETDomäne spezifischen Mutationen unterliegt, können die
SET-spezifischen DNA-Sonden zur Analyse von
Einzelstrang-Konformationspolymorphismen (single strand
conformation polymorphisms; SSCP; Gibbons et al., 1995)
eingesetzt werden.

Krebsarten, bei denen SET-spezifische DNA-Sonden als diagnostischer Marker verwendet werden können, sind Brustkrebs (*EZH1*; Friedman et al., 1994), Synovialksarkom (*SUV39H*; Geraghty et al.; 1993) und Leukämien.

Aufgrund der Kenntnis der Nukleotidsequenz der SETDomäne-Gene können die entsprechenden von der cDNASequenz abgeleiteten Proteine, die ebenfalls Gegenstand
der vorliegenden Erfindung sind, in rekombinanter Form
herstellt werden, indem die dafür kodierenden cDNAs in
geeignete Vektoren inseriert und in Wirtsorganismen
exprimiert werden. Die für die Produktion rekombinanter
Proteine verwendeten Techniken sind dem
Durchschnittsfachmann geläufig und können einschlägigen
Handbüchern (Sambrook, J., Fritsch, E.F. und
Maniatis, T., 1989, Cold Spring Harbor Laboratory
Press) entnommen werden.

Gegenstand der vorliegenden Erfindung sind somit in einem weiteren Aspekt rekombinante DNA-Moleküle, enthaltend die für EZH2, SUV39H oder EZH1 kodierende DNA sowie damit funktionell verbundene Expressionskontrollsequenzen, sowie die damit transformierten Wirtsorganismen.

Die erfindungsgemäßen rekombinanten Proteine können eingesetzt werden, um die Wechselwirkung von SETDomäne-Proteinen mit Chromatin bzw. mit anderen
Mitgliedern von Heterochromatinkomplexen zu
analysieren; ausgehend von den dabei erhaltenen
Erkenntnissen über die Wirkungsweise dieser Komplexe
werden die sich im einzelnen ergebenden Möglichkeiten
für das gezielte Eingreifen in die daran beteiligten
Mechanismen definiert und können für therapeutische
Anwendungen genutzt werden.

Untersuchungen, die der weiteren Analyse der Funktion der SET-Domäne dienen, werden z.B. durchgeführt, indem cDNAs, kodierend für humanes EZH2 bzw. SUV39H und versehen mit einem Epitop, gegen das Antikörper zur

Verfügung stehen, in vitro sowie in Gewebekulturen exprimiert werden. Nach Immunpräzipitation mit den jeweiligen epitopspezifischen Antikörpern kann festgestellt werden, ob EZH2 und SUV39H miteinander in vitro wechselwirken können und ob in vivo eine Komplexbildung zwischen EZH2 und/oder SUV39H mit weiteren Chromatinregulatoren stattfindet.

Es wurde bereits von anderen Autoren angenommen (DeCamillis et al., 1992; Rastelli et al., 1993; Orlando und Paro, 1993), daß eine Komplexbildung zwischen verschiedenen Mitgliedern von Heterochromatinproteinen wesentlich für deren Funktion ist. Aufgrund der Verfügbarkeit der erfindungsgemäßen SET-Domäne-Gene kann festgestellt werden, ob die SET-Region eine Domane darstellt, die aufgrund von Wechselwirkungen funktioniert, oder ob sie zum Entstehen multimerer heterochromatischer Komplexe beiträgt. Ebenso kann festgestellt werden, ob die SET-Domäne eine inhibitorische Funktion hat, ähnlich der aminoterminalen BTB-Domane verschiedener Chromatinregulatoren, einschließlich des GAGA-Faktors (Adams et al., 1992). Insgesamt erlauben die Analysen von Wechselwirkungen mit Epitop-versehenen EZH2- und SUV39H-Proteinen eine weitere Charakterisierung der Funktion der SET-Domäne. Dadurch werden Möglichkeiten eröffnet, gegen die deregulierte Aktivität vorzugehen, indem z.B. mittels gentherapeutischer Methoden dominant-negative Varianten der SET-Domäne-cDNA-Sequenzen in die Zelle eingeführt werden. Derartige Varianten werden z.B. erhalten, indem zunächst die funktionellen Domänen der SET-Proteine definiert werden, z.B. die für die DNA/Chromatin- oder die für die Protein/Protein-Wechselwirkung verantwortlichen Sequenzabschnitte, und indem dann die um die jeweilige(n) Domäne(n) oder um Abschnitte davon

verkürzten DNA-Sequenzen in der betroffenen Zelle zur Expression gebracht werden, um eine durch das intakte funktionelle Protein verursachte Deregulation der Proliferation zu kompetitieren.

Die Verfügbarkeit der erfindungsgemäßen cDNAs erlaubt ferner die Herstellung transgener Tiere, z.B. Mäuse, in denen SET-Domäne-Gene entweder überexprimiert werden können ("gain-of-function") oder in denen diese Gene ausgeschaltet werden können ("loss-of-function"); für letztere Analysen werden die korrespondierenden Tiersequenzen, insbesondere Maussequenzen, der erfindungsgemäßen Gene eingesetzt. Diese Mäuse sind ebenfalls Gegenstand der vorliegenden Erfindung.

Insbesondere gibt die "gain-of-function" - Analyse, bei der Allele der erfindungsgemäßen Gene in die Maus eingeführt werden, letztlich Aufschluß über die ursächliche Beteiligung von EZH2 und SUV39H an der chromatinabhängigen Förderung der Tumorentstehung. Für die "gain-of-function"- Analyse können die kompletten cDNA-Sequenzen von humanem EZH2 und SUV39H sowie deren mutierte Versionen, wie EZH1/B52 und MG-44, mittels Vektoren, die hohe Expressionsraten ermöglichen, z.B. Plasmiden mit dem humanen ß-Actin-Promotor, getrieben sowohl vom Enhancer der schweren Kette von Immunglobulinen (Eµ) als auch von Moloney-Virus-Enhancern (Mo-LTR). Kürzlich wurde gezeigt, daß die Eμ/Mo-LTR-abhängige Überexpression des bmi-Gens, welches gemeinsam mit EZH2 zur Pc-Gruppe negativer Chromatinregulatoren zählt, ausreicht in transgenen Mäusen Lymphome zu erzeugen (Alkema et al., 1995).

Indem in "loss-of-function"-Analysen die endogenen
Maus-Loci für Ezhl und Suv39h durch homologe
Rekombination in embryonalen Stammzellen unterbrochen

17

werden, kann bestimmt werden, ob der Verlust der in vivo-Genfunktion zu einer abnormalen Entwicklung der Maus führt.

Aufgrund dieser in vivo-Systeme kann die Wirkung von EZH2 und SUV39H bestätigt werden; diese Systeme dienen außerdem als Grundlage für Tiermodelle im Hinblick auf eine humane Gentherapie.

Der gentherapeutische Einsatz der erfindungsgemäßen DNA-Sequenzen oder davon abgeleiteter Sequenzen (z.B. komplementärer Antisenseoligonukleotide) erfolgt - je nachdem, ob die zu behandelnde Erkrankung auf eine Deregulation von Chromatin infolge des Fehlens der funktionellen Gensequenz, oder aber infolge einer Überexpression der entsprechenden Gene zurückzuführen ist - durch Einführung der funktionellen Gensequenz, durch Inhibierung der Genexpression, z.B. mit Hilfe von Antisenseoligonukleotiden, oder durch Einführung einer für eine dominant-negative Mutante kodierenden Sequenz. Die Einführung der jeweiligen DNA-Sequenzen in die Zelle kann mit Hilfe von Standardmethoden für die Transfektion höherer eukaryotischer Zellen erfolgen, zu denen der Gentransfer mittels viraler Vektoren (Retrovirus, Adenovirus, Adeno-assoziiiertes Virus) oder mittels nicht-viralen Systemen auf Basis der Rezeptor-vermittelten Endozytose zählen; Übersichten über gebräuchliche Methoden werden z.B. von Mitani und Caskey, 1993; Jolly, 1994; Vile und Russel, 1994; Tepper und Mule, 1994; Zatloukal et al., 1993, WO 93/07283 gegeben.

Für eine Inhibierung der Expression der erfindungsgemäßen Gene kommen auch niedermolekulare Substanzen in Betracht, die in die Transkriptionsmaschinerie eingreifen; nach Analyse der 5'-regulatorischen Region der Gene kann nach Substanzen gescreent werden, die die Wechselwirkung der jeweiligen Transkriptionsfaktoren mit dieser Region ganz oder teilweise blockieren, z.B. mit Hilfe der in der WO 92/13092 beschriebenen Methode.

Eine Inhibierung der deregulierten Proliferation kann auch am Genprodukt ansetzen, indem die entsprechenden Antikörper gegen das EZH2- oder SUV39H-Protein, vorzugsweise humane oder humanisierte Antikörper, therapeutisch eingesetzt werden. Die Herstellung solcher Antikörper erfolgt nach bekannten Methoden, wie sie z.B. von Malavsi und Albertini, 1992, oder von Rhein, 1993, beschrieben wurden.

Antikörper gegen EZH2 oder SUV39H, die therapeutisch oder diagnostisch verwendet werden können, sind ebenfalls Gegenstand der vorliegenden Erfindung.

Figurenübersicht

- Aminosäure-Sequenzvergleich zwischen EZH2 Fig. 1: und E(z)
- Aminosäure-Sequenzvergleich zwischen SUV39H Fig. 2: und Su(var)3-9
- Aberrante Transkripte von humanen SET-Domäne-Fig. 3: Genen
- Evolutionare Konservierung von SET-Domane-Fig. 4: Proteinen
- Fig. 5: Aminosaure-Übereinstimmung in der SET-Domäne
- DNA- und Aminosäure-Sequenz von EZH2 Fig. 6:
- DNA- und Aminosäure-Sequenz von SUV39H Fig. 7:
- partieller Sequenzvergleich zwischen den Fig. 8: cDNAs von humanem EZH2 und EZH1

Beispiel

a) Herstellung einer cDNA-Bibliothek

Es wurde eine humane B-Zell-spezifische cDNA-Bibliothek, wie von Bardwell und Treisman, 1994, beschrieben, hergestellt, indem Poly(A)+-RNA aus humanen BJA-B-Zellen isoliert, mittels Poly(dT)₁₅-Priming revers transkribiert und in doppelsträngige cDNA konvertiert wurde. Nach Zugabe eines EcoRI-Adapters der Sequenz 5' AATTCTCGAGCTCGTCGACA wurde die cDNA in die EcoRI-Stelle des Bakteriophagen gt10 ligiert. Die Propagierung und Amplifizierung der Bibliothek erfolgte in E.coli C600.

b) Herstellung von DNA-Sonden

Drosophila-DNA-Sonden, kodierend für die konservierten SET-Domänen von E(z) und Su(var)3-9, wurden auf der Grundlage der publizierten Drosophila-Sequenzen (Jones und Gelbart, 1993; Tschiersch et al., 1994) mittels Polymerasekettenreaktion (PCR) hergestellt: 1 μ g von Drosophila melanogaster-DNA (Clontech) wurde mit den beiden Primern E(z) 1910

- (5'ACTGAATTCGGCTGGGGCATCTTTCTTAAGG) und E(z) 2280
- (5' ACTCTAGACAATTTCCATTTCACGCTCTATG) einer PCR-Amplifikation (35 Zyklen zu 30 sek bei 94°C, 30 sek bei 55°C und 30 sek bei 72°C) unterworfen. Die entsprechende SET-Domäne-Sonde für Su(var)3-9 wurde von 10 ng Plasmid-DNA (Tschiersch et al., 1994; Klon M4) mit dem Primerpaar suvar.up
- (5' ATATAGTACTTCAAGTCCATTCAAAAGAGG) und suvar.dn
- (5' CCAGGTACCGTTGGTGCTGTTTAAGACCG) amplifiziert, wobei dieselben Zyklusbedingungen verwendet wurden. Die erhaltenen SET-Domäne-DNA-Fragmente wurden Gel-

PCT/EP96/01818 WO 96/35784

20

gereinigt und teilsequenziert, um die Richtigkeit der amplifizierten Sequenzen zu bestätigen.

c) Screenen der cDNA-Bibliothek

 5×10^5 Plaque-bildende Einheiten ("plaque forming units", pfu) wurden mit 5 ml Kultur des bakteriellen Wirtsstammes E.coli C600 (bei einer optischen Dichte OD600 von 0.5 in 10 mM MgSO4 suspendiert) bei 37°C 15 min lang inkubiert und dann auf eine große (200 mm x 200 mm) vorgewärmte LB-Schale ausgegossen. Nach Wachstum über Nacht bei 37°C wurden die Phagen auf einer Nylonmembran (GeneScreen) absorbiert. Die Membran wurde, die Seite mit den absorbierten Phagen nach oben schauend, 30 sek lang in Denaturierungslösung (1.5 M NaCl, 0.5 M NaOH) schwimmen gelassen, dann 60 sek in Denaturierungslösung eingetaucht und abschließend 5 min lang in 3 M NaCl, 0.5 M Tris pH 8, neutralisiert. Dann wurde die Membran kurz in 3xSSC gespült und die Phagen-DNA mittels UV-Vernetzung auf das Nylonfilter fixiert. Das Filter wurde 30 min lang bei 50°C in 30 ml Church-Puffer (1 % BSA, 1 mM EDTA und 0.5 M NaHPO₄, pH 7.2) prähybridisiert, anschließend wurden 2 x 106 cpm der radioaktiv markierten DNA-Sondenmischung (E(z)-SET und Su(var) 3-9-SET) zugegeben; die DNA-Sonden wurden durch Random-priming unter Verwendung des RediPrime-Kit (Amersham) hergestellt. Die Hybridisierung wurde über Nacht bei 50°C durchgeführt. Nach Entfernung der Hybridisierungslösung wurde das Filter 10 sek in 2xSSC, 1 % SDS bei Raumtemperatur gewaschen, anschließend wurde 10 sek lang bei 50°C gewaschen. Das Filter wurde in Saranwrap gewickelt und unter Verwendung einer Verstärkerfolie der Autoradiographie unterworfen.

Positive Phagenkolonien wurden auf der Originalplatte mittels Zuordnung des Autoradiogramms identifiziert und die entsprechenden Agarstückchen mit dem größeren Ende einer Pasteur-Pipette entfernt. Der Phagen-Pool wurde über Nacht bei 4° C in 1 ml SM-Puffer (5,8 g NaCl, 2 g MgSO₄-H₂O, 50 ml Tris pH 7.5, 5 ml 2 %ige Gelatine auf 1 l H₂O), enthaltend einige Tropfen CHCl₃, eluiert. Das Phagenlysat wurde für eine zweite und dritte Screeningrunde wieder ausplattiert um einzelne, gut isolierte positive Plaques (20 bis 100 Plaques pro Platte in der dritten Runde) zu erhalten.

d) Sequenzanalyse

Die cDNA-Inserts von rekombinanten Phagen wurden in den Polylinker von pBluescript KS (Stratagene) subkloniert und auf einem automatischen Sequenzer (Applied Biosystems) unter Verwendung der Didesoxy-Methode sequenziert. Die komplette Sequenz von wenigstens zwei unabhängigen Isolaten pro erhaltenen Gen wurde mittels "Primer walking" ermittelt. Die Sequenzen wurden mit dem GCG-Software-Paket (University of Wisconsin) analysiert, die Untersuchung auf Homologie wurde mit dem "Blast and fasta" oder "tfasta" Netzwerkservice durchgeführt. Die kompletten Sequenzen von EZH2 und SUV39H sind in Fig. 6 und 7 dargestellt.

Literatur

Adams et al., 1992, Genes & Dev. 6, 1589-1607.

Alkema et al., 1995, Nature 374, 724-727.

Allshire et al., 1994, Cell 76, 157-169.

Bardwell und Treisman, 1994, Genes & Dev. 8, 1644-1677.

Brunk et al., 1991, Nature 353, 351-355.

Buck und Shore, 1995, Genes & Dev. 9, 370-384.

Cleary, 1991, Cell 66, 619-622.

Counter et al., 1992, Embo J. 11, 1921-1928.

DeCamillis et al., 1992, Genes & Dev. 6, 223-232.

Eissenberg et al., 1992, Genetics 131, 345-352.

Friedman et al., 1994, Cancer Research 54, 6374-6382.

Garzino et al., 1992, Embo J. 11, 4471-4479.

Geraghty et al., 1993, Genomics 16, 440-446.

Gibbons et al., 1995, Cell 80, 837-845.

Gu et al., 1992, Cell 71, 701-708.

Haupt et al., 1991, Cell 65, 753-763.

Jolly, D., 1994, Cancer Gene Therapy 1, 51.

Jones und Gelbart, 1993, MCB 13 (10), 6357-6366.

Kennedy et al., 1995, Cell 80, 485-496.

Locke et al., 1988, Genetics 120, 181-198.

Malavsi, F. und Albertini, A., 1992, TIBTECH 10, 267-269.

Messmer et al., 1992, Genes & Dev. 6, 1241-1254.

Milner und Campbell, 1993, Biochem. J. 290, 811-818.

Mitani, K. und Caskey, C.T., 1993, Trends in Biotechnology 11, 162-166.

Nomura et al., 1994, Unpublished. GeneBank accession number: D31891.

Orlando und Paro, 1993, Cell 75, 1187-1198.

Pardue, 1991, Cell 66, 427-431.

Rastelli et al., 1993 Embo J. 12, 1513-1522.

Renauld et al., 1993, Genes & Dev. 7, 1133-1145.

Reuter und Spierer, 1992, BioEssays 14, 605-612.

Rhein, R., 1993, The Journal of NIH Res. 5, 40-46.

Smouse und Perrimon, 1990, Dev. Biol. 139, 169-185.

Solomon et al., 1991, Science 254, 1153-1160.

Tepper, R.I. und Mule, J.J., 1994, Human Gene Therapy
5, 153.

Tkachuk et al., 1992, Cell 71, 691-700.

Tschiersch et al., 1994, Embo J. 13 (16), 3822-3831.

Vile, R. und Russel S., 1994, Gene Therapy 1, 88.

Zatloukal, K., Schmidt, W., Cotten, M., Wagner, E.,

Stingl, G. und Birnstiel, M.L., 1993, Gene 135,

199.

Patentansprüche

- 1. DNA-Moleküle, enthaltend eine für ein Chromatinregulator-Protein, das eine SET-Domäne aufweist, kodierende Sequenz oder eine Teilsequenz davon, dadurch gekennzeichnet, daß sie die in Fig. 6, Fig. 7 oder Fig. 8 dargestellte Nukleotidsequenz, kodierend für EZH2, SUV39H oder EZH1, einschließlich ihrer degenerierten Varianten sowie Mutanten davon, enthalten.
- DNA-Molekül nach Anspruch 1, dadurch gekennzeichnet, daß es eine cDNA ist.
- 3. DNA-Molekül nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß es humanen Ursprungs ist.
- 4. cDNA nach Anspruch 3 der Bezeichnung EZH2.
- 5. cDNA nach Anspruch 3 der Bezeichnung SUV39H.
 - 6. DNA-Molekül nach Anspruch 4, dadurch gekennzeichnet, daß es die für die SET-Domäne von EZH2 kodierende Region oder einen Abschnitt davon enthält.
 - 7. DNA-Molekül nach Anspruch 5, dadurch gekennzeichnet, daß es die für die SET-Domäne SUV39H kodierende Region oder einen Abschnitt davon enthält.
 - 8. DNA-Molekül nach Anspruch 4, dadurch daß es für eine dominant-negative Mutante von *EZH2* kodiert.

25

9. DNA-Molekül nach Anspruch 4, dadurch daß es für eine dominant-negative Mutante von SUV39H kodiert.

- 10. Rekombinantes DNA-Molekül, enthaltend eine in Anspruch 2 definierte cDNA, funktionell verbunden mit Expressionskontrollsequenzen, zur Expression in prokaryotischen oder eukaryotischen Wirtsorganismen.
- 11. Prokaryotische oder eukaryotische
 Wirtsorganismen, transformiert mit rekombinanter
 DNA nach Anspruch 10.
- 12. Rekombinantes humanes Chromatinregulator-Protein *EZH2* oder ein Abschnitt davon, erhältlich durch Expression einer in Anspruch 4 definierten cDNA.
- 13. Rekombinantes humanes Chromatinregulator-Protein SUV39H oder ein Abschnitt davon, erhältlich durch Expression einer in Anspruch 5 definierten cDNA.
- 14. Antikörper gegen EZH2.
- 15. Antikörper gegen SUV39H.
- 16. Antisense (desoxy) ribonukleotide mit

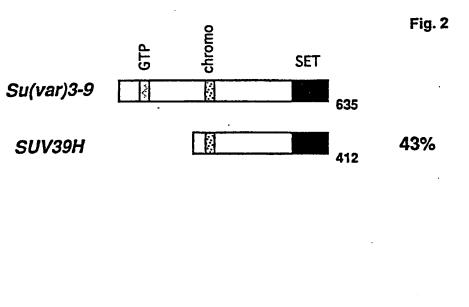
 Komplementarität zu einer Teilsequenz einer in

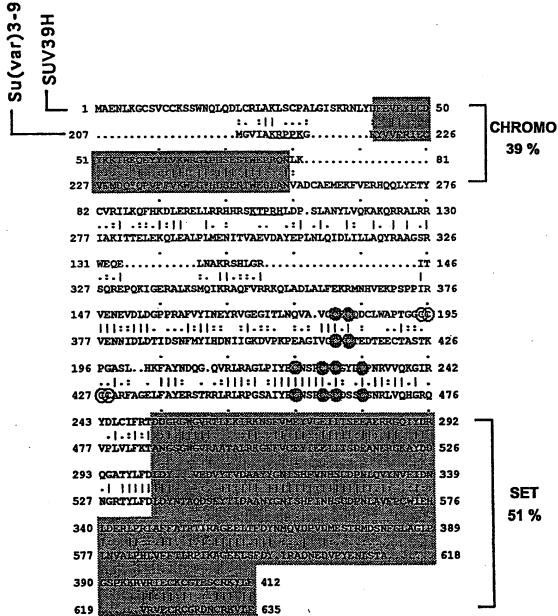
 Anspruch 1 definierten DNA.
- 17. DNA-Molekül, kodierend für die SET-Domäne eines Chromatinregulator-Gens, bzw. für einen Abschnitt davon, zur Behandlung und Diagnose von Erkrankungen des Menschen, die auf eine

Deregulation von Chromatinregulator-Genen, die eine SET-Domäne aufweisen, zurückzuführen sind.

- 18. DNA-Molekül nach Anspruch 6 oder 7 zur
 Behandlung und Diagnose von Erkrankungen des
 Menschen, die auf eine Deregulation von
 Chromatinregulator-Genen, die eine SET-Domāne
 aufweisen, zurückzuführen sind.
- 19. Antikörper nach Anspruch 14 oder 15 zur Behandlung und Diagnose von Erkrankungen des Menschen, die auf eine Deregulation von Chromatinregulator-Genen, die eine SET-Domäne aufweisen, zurückzuführen sind.
- 20. Transgene Maus, enthaltend ein Transgen für die Expression eines Chromatinregulator-Gens, das eine SET-Domäne aufweist, oder einer mutierten Version eines solchen Proteins.
- 21. Knock-out Maus, erhältlich aus embryonalen Stammzellen, in denen die endogenen Maus-Loci für Ezhl und Suv39h durch homologe Rekombination unterbrochen wurden.
 - 22. Verfahren zum Identifizieren von SäugetierChromatinregulator-Genen, die eine SET-Domäne
 aufweisen, oder mutierten Versionen davon,
 dadurch gekennzeichnet, daß man Säugetier-cDNAoder genomische DNA-Bibliotheken unter
 Bedingungen niedriger Stringenz mit einem DNAMolekül, kodierend für die SET-Domäne oder einen
 Abschnitt davon, hybridisiert.

		•	
EZH2 E(z)		MGQTGKKSEKGPVCWRKRVKSEYMRLRQLKRFRRADEVKSMFSSNRQKIL 50!::::::::::::::::::::::::::::::::	Fig. 1
		ERTEILNQEWKORRIQPVHILTSVSSLRGTRECSVTSDLDFPTQVIPL 98	
		: :	
		KTLNAVASVPIMYSWSPLOONFMVEDETVLHNIPYMGDEVLDQDGTFIEE 148	
		CVINAVTPIPTMYTWAPTQONFMVEDETVLHNIPYMGDEVLDKDGKFIEE 141	
٠		LIKNYDGKVHGDRECGFINDEIFVELVNAL	
			•
		GQYNDDDDDDDDDDDDEEREEKQKDLEDH 206 : : :. . :. . TEPLAKSKQGEDDGVVDVDADCESPMKLEKTESKGDLTDVEKKETEEPVE 241	
	207	RDDKESRPPRKFPSDKIFEAISSMFPDKGTAEELKEKYKELTE 249	
		: :: : :	
	250	QOLPGALPPECTPNIDGPNAKSVOREQSLHSFHTLECKECKYDGTLHPF 299	
	292	: : . . . :	
	300	HATPNTYKRKNTETALDNKE POPULEGAKEFAAALTAERIKTPP 347	
	341	OGHAGPNIOKRRYPELKPFAERENSOMLIDGMKEKLAADSKTPP 386	
		KRPGGRRGRLPNNSSRPSTPTINVLESKDTDSDREAGTETGGENNDKEE 397	
•		IDSCNEASSEDSNDSNSQFSNKDFNH 412	
		EEKKDET.SSSSEANSRCQTPIKMKPNIEPPENVENSGAEASMFRVLIGT 446 :. : : . : . ENSKDNGLTVNSAAVAEINSIMAGMMNITSTQCV.WTGADQALYRVLHKV 461	
		YYDNE AIARLIGTKY ROVYEFRVKESSIIAPAPAEDVD <u>TPPRKKKRK</u> H 496	
	462	YLKNYAIAHNMLTKTERQVYEFAQKEDAEFSFEDLRQDFTPPRKKKKKQ 511	
	497	RIMAAH RKIOLKKOGSSNHVYNYOF DHPROPOSS VIAONEEKF 546 —	١
	512	-	C-reich
	547	CONTROL CANDING CONTROL CONTRO	75%
	562	OCESTO ONREPCO CKACO TROO YLAVRE OPDE DAGS. ADOFKL 610 —	ļ
	597	KNVS KNCSIORGSKKHLLLAPSDVASNGIE IKDEVOKNERISEKGEEII 646 —	1
	611	:- -:: - :	
		SODEADRUGKVYDKYMCSELENLNNDEVVDATERGNKEREANHSVNPNOV 696	SET
		SQUEADRICKYTHY MUSELEN LANDE VYDATRICKKI KIRUKANIH SINENCY 710	88%
	697	7 AKYMYNGBERIGTETAKKA BYTGEELE-FOZRYSQADALK XVGLEREMETEP 746	
	711	BRUMBUTCHERIGTFARRATOPICETEFFOTRYSPTEOLEGYGERENERY 760 —	J





2/11

Fig. 3

ABERRANTE TRANSKRIPTE

НВХ	EZH1	Н6ЕЛЛЅ	<i>C9a</i>	~
ALL-1				KG-1
	B52	MG-44	~	
TRANSLOKATIONEN LEUKÄMIE	INSTABILE REGION BRUSTKREBS (BRCA1)	TRANSLOKATIONEN SYNOVIALSARKOM (OATL1)	(MHC III) AUTOIMMÜNERKRANKUNGEN	MYELOIDE TUMOREN
11923	17q21	7 Xp11	6p21	6

Fig. 4
SET PROTEIN FAMILIE

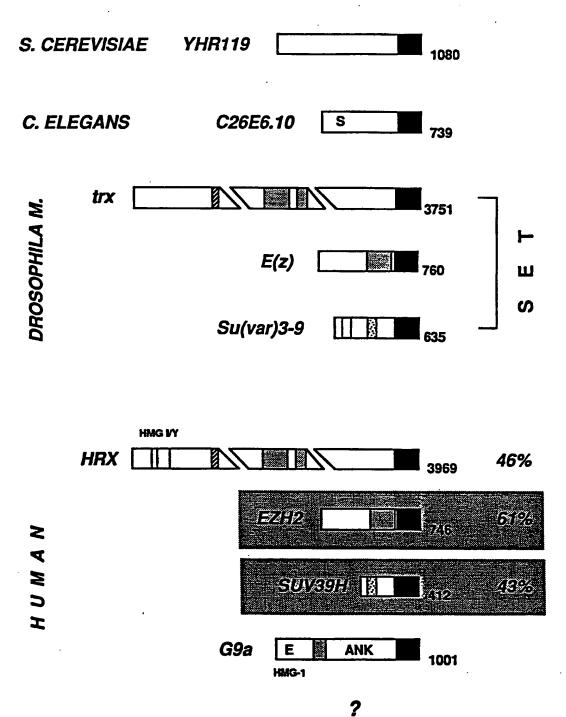


Fig. 5

P (-1	CDTAGGGGTSTST	VEC 3 OWNERT	C#WCEET#CO	DE SONO CELLES	DKYMCSFL	E0
E(z) EZH2					DKYMCSFL	30
HRX	SPINSPEIEC	KDNIDACEMI	TEVACMUEDO	TOTOURERVY	DSKGIG.CYM	
trx	60000000		G00004 5000 900	-942 SSM SS6	DSRGIG.CYM	
C26	464056500		C00000 3000 9000 T	LVAEEREKAY		
YHR	500000000		5506080 6007 8000	PVAEMREKRY		
IIIK	SATHUMGLIA	DOSIARCENT	TELVOERTRO	FVAERIKERRE	TWG16221F	
Su3-9	ANGSONGUDA	ATAI.REGEEU	CENTERTATE	DEANERGKAY	DDNG PEVI.	
SUV39H				EEAERRGOIY		
G9a				AEADV.		
KG-1				DFADKEGL.		
	2	,	O###-##-12	21,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,		
		/				
	485 488		40000 4000 ABANAN	*****		
E(z)					MVTGDH	100
EZH2	ININ	NDFVVDATRK	GNKIRFANHS	VNPNCYAKVM	MVNGDH	
					•	
HRX				CEPNCYSRVI		
trx	222 593	909000	MCCCCOQ XXXX MCCCCCC	CEPNCYSKVV		
C26	276 826	\$20000	200000 0000000000000000000000000000000	COPNCYAKVL		
YHR	ERVD	ENTVIDATKK	EGIAREINHC	CDPNCTAKII	KVGGRR	
Su3-9		CEVETERALIV				
SUV39H				CDPNLAVFPC CDPNLQVYNV		
50 v 39n G9a	EDIENK DC	EVYCTREBYY	ENT COUTURE	CDENLIPVRV	E I DNI DE KILP	
KG-1				CSPNLFVQNV		
	W.1-27.17.	Λ	####C. CENTRICE	COMMENT VOIVV	r voimble r	
		4				
	**** *** *** *** ***	AND TO A STATE OF THE STATE OF		***		
E(z)				KFVĞI		150
E (z) EZH2	RIGIFAKRAI	OTGEELFFDY	RYSQADAL	KYVGI	EREMEIP*	150
	RIGIFAKRAI HIVIFAMRKI	QTGEELFFDY YRGEELTYDY	RYSQADAL	KYVĞI SNKLPCNCGA	EREMEIP* KKCRKFLN*	150
EZH2 HRX trx	RIGIFAKRAI HIVIFAMRKI HIIIFAVRRI	OTGEELFER YRGEELTYRY VOGEELTYRY	RYSQADALKFPIE.DAKFPFE.D.	KYVĞI SNKLPCNCGA .EKIPCSCGS	EREMEIP* KKCRKFLN* KRCRKYLN*	150
EZH2 HRX trx C26	RIGIFAKRAI HIVIFAMRKI HIIIFAVRRI REVIXSRTII	OTGEELFEDY YRGEELTYDY VOGEELTYDY KKGEELTYDY	RYSQADALKFPIE.DAKFPFE.DKFPIE	KYVĞI SNKLPCNCĞA .EKIPCSCĞS DDKIDCLCĞA	EREMEIP* KKCRKFIN* KRCRKYIN*	150
HRX trx C26 YHR	RIGIFAKRAJ HIVIFAKRIJ HILIFAVRRIJ RIVIVARIJA RIVIVALROJ	OTGEELFFDY YRGEELTYDY VOGEELTYDY KKGEELTYDY AASBEELTYDY	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA	EREMEIP* KKCRKFIN* KRCRKYIN*	150
EZH2 HRX trx C26 YHR Su3-9	RIGIFAKRAJ HIVIFAMRKI HIVIFAVRIJ KIVIFALROJ HIVFFILROJ	QTGEELFFDY YRGEELTYDY VQGEELTYDY KKGEELTYDY AASBELTYDY KAGEELSFDY	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN*	150
HRX trx C26 YHR Su3-9 SUV39H	REGIEARRAJ HEVIEAMRKI HELIEAVREJ REVIESETIJ REVIESETIROJ HEVFETEROJ REAFFATROJ	OTGEELFFDI YRGEELTYDY VOCEELTYDY KKGEELTYDY AASPELTYDY KAGEELSFDY RAGEELTFDY	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR	150
HRX trx C26 YHR Su3-9 SUV39H G9a	RIGIEAKRAJ HIVIEAMRKI HIVIESKIJ RIVIESKIJ RIVIETIKOJ HIVFETIKOJ RIAFFATRIJ RIAFFATRIJ	QTGEELFFDX YRGEELTYDX VQGEELTYDX KKGEELTYDX AASEELTYDX KAGEELSFDX RAGEELJFDX RTGEELGFDY	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL SKYFTCQCS	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150
HRX trx C26 YHR Su3-9 SUV39H	RIGIEAKRAJ HIVIEAMRKI HIVIESKIJ RIVIESKIJ RIVIETIKOJ HIVFETIKOJ RIAFFATRIJ RIAFFATRIJ	QTGEELFFDX YRGEELTYDX VQGEELTYDX KKGEELTYDX AASEELTYDX KAGEELSFDX RAGEELJFDX RTGEELGFDY	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150
HRX trx C26 YHR Su3-9 SUV39H G9a	RIGIEAKRAJ HIVIEAMRKI HIVIESKIJ RIVIESKIJ RIVIETIKOJ HIVFETIKOJ RIAFFATRIJ RIAFFATRIJ	QTGEELFFDX YRGEELTYDX VQGEELTYDX KKGEELTYDX AASEELTYDX KAGEELSFDX RAGEELJFDX RTGEELGFDY	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL SKYFTCQCS	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150
HRX trx C26 YHR Su3-9 SUV39H G9a	RIGIEAKRAJ HIVIEAMRKI HIVIESKIJ RIVIESKIJ RIVIETIKOJ HIVFETIKOJ RIAFFATRIJ RIAFFATRIJ	QTGEELFFDX YRGEELTYDY VQGEELTYDY KKGEELTYDY AASEELTYDY KAGEELSFDY RAGEELTFDY RTGEELGFDY	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL SKYFTCQCS	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150
HRX trx C26 YHR Su3-9 SUV39H G9a	RIGIEAKRAJ HIVIEAMRKI HIVIESKIJ RIVIESKIJ RIVIETIKOJ HIVFETIKOJ RIAFFATRIJ RIAFFATRIJ	QTGEELFFDX YRGEELTYDY VQGEELTYDY KKGEELTYDY AASEELTYDY KAGEELSFDY RAGEELTFDY RTGEELGFDY	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL SKYFTCQCS	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150
HRX trx C26 YHR Su3-9 SUV39H G9a KG-1	RIGIEAKRAJ HIVIEAMRKI HIVIESKIJ RIVIESKIJ RIVIETIKOJ HIVFETIKOJ RIAFFATRIJ RIAFFATRIJ	QTGEELFFDX YRGEELTYDY VQGEELTYDY KKGEELTYDY AASEELTYDY KAGEELSFDY RAGEELTFDY RTGEELGFDY	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL SKYFTCQCS	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150
EZH2 HRX trx C26 YHR Su3-9 SUV39H G9a KG-1 E(z) EZH2	RIGIEAKRAJ HIVIEAMRKI HIVIESKIJ RIVIESKIJ RIVIETIKOJ HIVFETIKOJ RIAFFATRIJ RIAFFATRIJ	QTGEELFFDX YRGEELTYDY VQGEELTYDY KKGEELTYDY AASEELTYDY KAGEELSFDY RAGEELTFDY RTGEELGFDY	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL SKYFTCQCS	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150
HRX trx C26 YHR Su3-9 SUV39H G9a KG-1	RIGIEAKRAJ HIVIEAMRKI HIVIESKIJ RIVIESKIJ RIVIETIKOJ HIVFETIKOJ RIAFFATRIJ RIAFFATRIJ	QTGEELFFDX YRGEELTYDY VQGEELTYDY KKGEELTYDY AASEELTYDY KAGEELSFDY RAGEELTFDY RTGEELGFDY	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL SKYFTCQCS	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150
HRX trx C26 YHR Su3-9 SUV39H G9a KG-1 E(z) EZH2 HRX trx	RIGIEAKRAJ HIVIEAMRKI HIVIESKIJ RIVIESKIJ RIVIETIKOJ HIVFETIKOJ RIAFFATRIJ RIAFFATRIJ	QTGEELFFDX YRGEELTYDY VQGEELTYDY KKGEELTYDY AASEELTYDY KAGEELSFDY RAGEELTFDY RTGEELGFDY	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL SKYFTCQCS	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150
HRX trx C26 YHR Su3-9 SUV39H G9a KG-1 E(z) EZH2 HRX trx 26	RIGIEAKRAJ HIVIEAMRKI HIVIESKIJ RIVIESKIJ RIVIETIKOJ HIVFETIKOJ RIAFFATRIJ RIAFFATRIJ	QTGEELFFDX YRGEELTYDY VQGEELTYDY KKGEELTYDY AASEELTYDY KAGEELSFDY RAGEELTFDY RTGEELGFDY	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL SKYFTCQCS	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150
HRX trx C26 YHR Su3-9 SUV39H G9a KG-1 E(z) EZH2 HRX trx	RIGIEAKRAJ HIVIEAMRKI HIVIESKIJ RIVIESKIJ RIVIETIKOJ HIVFETIKOJ RIAFFATRIJ RIAFFATRIJ	QTGEELFFDX YRGEELTYDY VQGEELTYDY KKGEELTYDY AASEELTYDY KAGEELSFDY RAGEELTFDY RTGEELGFDY	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL SKYFTCQCS	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150
EZH2 HRX trx C26 YHR Su3-9 SUV39H G9a KG-1 E(z) EZH2 HRX trx 26 YHR	REGIFARRAJ ELVIFAMRKI REVIFSRETI REVIFALROJ HLVFETLROJ REAFFATREJ REAFFASKOJ WYAFFASKOJ	QTGEELFFD: YRGEELTYDY VQGEELTYDY KKGEELTYDY AASBELTYDY KAGEELSFD: RAGEELTFD: RTGEELGFPY RAGTELTWD:	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL SKYFTCQCS	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150
HRX trx C26 YHR Su3-9 SUV39H G9a KG-1 E(z) EZH2 HRX trx 26	REGIFARRAJ ELVIFAMRKI HILIFAVRRI RIVIESRIJ HIVFETIROJ RIAFFATRIJ RIAFFASROJ WYAFFASROJ	QTGEELFFD: YRGEELTYDY VQGEELTYDY KKGEELTYDY AASBELTYDY KAGEELSFD: RAGEELTFD: RTGEELGFD: RAGTELTWD:	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL SKYFTCQCS	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150
EZH2 HRX trx C26 YHR Su3-9 SUV39H G9a KG-1 E(z) EZH2 HRX trx 26 YHR Su3-9	REGIFARRAJ ELVIFAMRKI HILIFAVRRI RIVIESRIJ HIVFETIROJ RIAFFATRIJ RIAFFASROJ WYAFFASROJ VRVECRCGRD VRVECRCGRD VRIECKCGTE	QTGEELFFD: YRGEELTYDY VQGEELTYDY KKGEELTYDY AASBELTYDY KAGEELSFD: RAGEELTFD: RTGEELGFD: RAGTELTWD:	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK NYEVGSVE	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL SKYFTCQCS	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150
EZH2 HRX trx C26 YHR Su3-9 SUV39H G9a KG-1 E(z) EZH2 HRX trx 26 YHR Su3-9 SUV39H	REGIFARRAJ ELVIFAMRKI HILIFAVRRI RIVIESRIJ HIVFETIROJ RIAFFATRIJ RIAFFASROJ WYAFFASROJ VRVECRCGRD VRVECRCGRD VRIECKCGTE	QTGEELFFDT YRGEELTYDY VQGEELTYDY KKGEELTYDY RASEELTYDY RASEELTYDY RAGEELTFDT RTGEELGFDY RAGTELTWDY NCRKVLF*	RYSQADALKFPIE.DAKFPFE.DKFPIEKFEREKDD .IRADNEDVP NMQVDPVDME GDRFWDIK NYEVGSVE	KYVGI SNKLPCNCGA .EKIPCSCGS DDKIDCLCGA EERLPCLCGA YENLSTA STRMDSNFGL SKYFTCQCS	EREMEIP* KKCRKFLN* KRCRKYLN* KTCRGYLN* PNCKGFLN* AGLPGSPKKR EKCKHSAEAI	150

Fig. 6/1

EZH2 Länge: 2600bp (kodierend: 90 - 2330)

1	AGGCAGTGGAGCCCCGGCGGCGCGCGGCGCGCGCGCGGGAACAACG	60
61	CGAGTCGGCGCGGGACGAAGAATAATCATGGGCCAGACTGGGAAGAAATCTGAGAAGG M G Q T G K K S E K G	120
121	GACCAGTTTGTTGGCGGAAGCGTGTAAAATCAGAGTACATGCGACTGAGACAGCTCAAGA P V C W R K R V K S E Y M R L R Q L K R	180
181	GGTTCAGACGAGCTGATGAAGTAAAGAGTATGTTTAGTTCCAATCGTCAGAAAATTTTGG F R R A D E V K S M F S S N R O K I L E	240
241	AAAGAACGGAAATCTTAAACCAAGAATGGAAACAGCGAAGGATACAGCCTGTGCACATCC	300
	RTEILNQEWKQRRIQPVHIL	
	· · · · ·	
301	TGACTTCTGTGAGCTCATTGCGCGGGACTAGGGAGTGTTCGGTGACCAGTGACTTGGATT	360
	TSVSSLRGTRECSVTSDLDF	
361	TTCCAACACAAGTCATCCCATTAAAGACTCTGAATGCAGTTGCTTCAGTACCCATAATGT	420
	PTQVIPLKTLNAVASVPIMY	
421	ATTCTTGGTCTCCCCTACAGCAGAATTTTATGGTGGAAGATGAAACTGTTTTACATAACA	480
	SWSPLQQNFMVEDETVLHNI	
481	TTCCTTATATGGGAGATGAAGTTTTAGATCAGGATGGTACTTTCATTGAAGAACTAATAA	540
101	PYMGDEVLDQDGTFIEELIK	340
541	AAAATTATGATGGGAAAGTACACGGGGATAGAGAATGTGGGTTTATAAATGATGAAATTT	600
	NYDGKVHGDRECGFINDEIF	
601		660
901	TTGTGGAGTTGGTGAATGCCCTTGGTCAATATAATGATGATGACGATGATGATGATGAGGAG V E L V N A L G Q Y N D D D D D D G D	660
661	ACGATCCTGAAGAAAGAGAAAAGCAGAAAGATCTGGAGGATCACCGAGATGATAAAG	720
	DPEEREEKQKDLEDHRDDKE	
721	AAAGCCGCCCACCTCGGAAATTTCCTTCTGATAAAATTTTTTGAAGCCATTTCCTCAATGT S R P P R K F P S D K I F E A I S S M F	780
781	TTCCAGATAAGGGCACAGCAGAAGAACTAAAGGAAAATATAAAGAACTCACCGAACAGC	840
	P D K G T A E E L K E K Y K E L T E Q Q	
841	AGCTCCCAGGCGCACTTCCTCCTGAATGTACCCCCAACATAGATGGACCAAATGCTAAAT	900
	LPGALPPECTPNIDGPNAKS	
901	CTGTTCAGAGAGAGCAAAGCTTACACTCCTTTCATACGCTTTTCTGTAGGCGATGTTTTA	960
301	V Q R E Q S L H S F H T L F C R R C F K	
961	AATATGACTGCTTCCTACATCCTTTTCATGCAACACCCAACACTTATAAGCGGAAGAACA	1020
	YDCFLHPFHATPNTYKRKNT	
.021	CAGAAACAGCTCTAGACAAACCATGTGGACCACAGTGTTACCAGCATTTGGACGGAG	
	ETALDNKPCGPQCYQHLEGA	
081	CANAGGAGTTTGCTGCTCTCACCGCTGAGCGGATAAGACCCCCACCAAAACGTCCAG	1140

Fig. 6/2

																	•			•	
1141	GAGG	ccc	CJG	מממ	acc:	NCG(- -	TCC	יממי	י. ממד	יאכי	PAC	ממח	ccc	CAG	CAC	הכר	cac	<u>ሮሽ ሞ</u>	TA	1200
TTAT																					1200
	G	ĸ	K	ĸ	G	ĸ	ь	ħ	. N	N	S	S	ĸ	P	5	T	2	1	1	N	
			•				•			•			•				•			•	
1201	ATGT	GCT	GGA	ATC	AAA	GA.	TAC	AGA	CAG:	TGA:	rag(GGA.	AGC	AGG	GAC	TGA	AAC	GGG	GGG	AG	1260
	v	L	E	S	K	D	T	D	S	D	R	E	A	G	T	E	T	G	G	E	
	-	_		_		_		_	_	_		_		_	_	_	_	_	_	_	
1061		~~ ~	- -	m > > ·		. ~ > :	•	. ~ .	~~ ~	~	. ~ > 1		•	mmc	~~~	~m~	~~~			•	1220
1261																					1320
	N	N	D	K	E	E	E	E	K	K	D	E	T	S	S	S	S	E	A	N	
						_			_												
1321	ATTC	ጥርር	CTC	ሞሮል	מממ	ACC:	מאמ		CAT	CAA	acc:	מממ	тат	TGA	ACC	TCC	TCA	CAA	TCT)	GG	1380
1321					T																2000
	5	K	C	¥	1	-	1	v	23	v	F	14	1	E	E	E	£	T.	٧	-	
			•				•			•			•				•			•	
1381	agtg	GAG	TGG	TGC	TGA	AGC	CTC	AAT	GTT	TAG	agt	CCT	CAT	TGG	CAC	TTA	CTA	TGA	CAA'	TT	1440
	W	S	G	A	E	A	S	M	F	R	V	L	I	G	T	Y	Y	D	N	F	
							_			_			_				_			_	
1441	TCTG	-	~~ T	TCC	መክ ለ	~ ₩₩		TCC	CAC	רא אי	N N C	יייי ג	ሞአር	202	CCT	CTA	· ጥር አ	CTT.	ምአ ር፡	D.C	1500
TAAT															_						1300
	C	A	Ţ	A	R	Į,	Ţ	G	T	K	T	C	K	Q	V	I	E	E	K	V	
			•				•			•			•				•			•	
1501	TCAA	AGA	ATC	TAG	CAT	CAT	agc	TCC.	AGC	TCC	CCC.	TGA	GGA	TGT	GGA	TAC	TCC	TCC	aag	GA	1560
	K	E	S	S	I	I	A	P	A	P	A	E	D	V	D	T	P	P	R	K	
			_				_						_				_			_	
1561	AAAA	CAR	CAC	.caa	مخم	rra	- CTT	CTC	ccc	TGC		СТС	CVG	444	СЪТ	262	CCT	C A A	222	22	1620
1301							_													•	+020
	K	K	K	V	H	ĸ	T	**	A	A	n	C	K	v	_	¥	L	v	v	U	
			•				•			•			•				•			•	
1621	ACGG	CTC	CTC	TAA	CCA	TGT	TTA	CAA	CTA	TCA	ACC	CTG	TGA	TCA	TCC	ACG	GCA	GCC	TTG	TG	1680
	G	S	S	N	H	v	Y	N	Y	Q	P	C	D	H	P	R	Q	P	C	D	
			_														_				
1681	ACAG	TTC	·CTC	ccc	ተ ጥር	TCT	- Gat	AGC	ACA	AAA	ттт	ተጥር	TGA	AAA	CTT	ም ተር	TCA	ATG	TAC	TT	1740
1001					c										-						
	3	9	C	r	C	٧	-	A	¥	14	-	C		K	_	•	~	C	3	3	
			•				•			•			•				•			•	
1741	CAGA	GTG	TCA	LAAA	CCG	CTT	TCC	GGG	ATG	CCG	CTG	CAA	AGC	ACA	GTG	CAA	CAC	CAA	GCA	GT	1800
	E	С	Q	N	R	F	P	G	C	R	C	K	A	Q	C	N	T	K	Ø.	С	
1801	GCCC	TTE	·CTZ	CCT	GGC	TGT	CCG	AGA	GTG	TGA	ccc	TGA	CCT	CTG	TCT	TAC	ттс	TGG	AĠC	CG	1860
7007					A																
	-	C	1	ш	A	•	A		_		•	•		·	-	•	·	G	A	_	
				•			•			•							•			•	
1861	CTGA															-					1920
	D	H	W	D	S	K	N	V	S	C	K	N	C	S	I	Q	R	G	S	K	
							•			•				,			•			-	
1921	AAA	AGCI	ATC:	TTA	'GCT	GGC	ACC	ATC	TGA	CGT	GGC	AGG	CTG	GGG	GAI	TTT	TAT	CAA	AGA	TC	1980
	K	Ħ	T.	T.	L	A	P	S	D	v	A	G	W	G	I	F	I	K	D	P	
			_	_	_		-	_	_	•		_	•••	_	_	_	_		_		
				•			•			•							•			•	
1981																					2040
	v	Q	K	N	E	F	I	S	E	Y	С	G	E	I	I	S	Q	D	E	A	
			٠.							•							•				
2041	CTG	CAC	2220	:AC	CAR	AGT	GTA	TGA	TAR	ATA	CAT	GTO	CAC	CT	rtci	GTI	CA	CTT	GAA	CA	2100
2012					K																
	U	K		G	A	•	-	U	K	•	F.7	•	3	-		E	74	-	74	14	
			•	•			•			•			•	•			•			•	
2101	ATG	ATT:	TTG:	rgg:	:GG/	LTGC	:AAC	200	CA	/GGC	TA	CA	LAA!	rtco	STT?	MGC	:AA:	/ICI	LTTC	EG	2160
	D	F	v	v	D	A	T	R	K	G	N	K	I	R	F	A	N	Ħ	S	v	
	_	-	-		_		_	-								-	_			_	
2161	m221			ا √سہ •	دنهماذ	\ T	- •	ابت لا _ا	n K dad		التان ا	יביי	, ,	تىلىت.	1 TY-1	\ \ \ \	יג בי	PAC	י ע קריב	-	2220
4101																					2220
	N	P	N	C	Y	A	K	V	M	M	V	N	G	D	H	R	I	G	1	F	
				•			•			•				•			•			•	
2221	TTG	CA	AGA	GAG	CAT	rccz	IGA (TG	CG/	AAGI	\GC1	GT	TT:	rtg:	ATT	ACAC	SAT	ACAC	SCC2	AGG	2280
					Ŧ																

PCT/EP96/01818

Fig. 6/3

2281	CTGATGCCCTGAAGTATGTCGGCATCGAAAGAGAAATGGAAATCCCTTGACATCTGCTAC	2340
	DALKYVGIEREMEIP*	
	•	
2341	CTCCTCCCCTCTCTGAAACAGCTGCCTTAGCTTCAGGAACCTCGAGTACTGTGGGCAA	2400
	• • • • • • • • • • • • • • • • • • • •	
2401	TTTAGAAAAAGAACATGCAGTTTGAAATTCTGAATTTGCAAAGTACTGTAAGAATAATTT	2460
	• • • • • • • • • • • • • • • • • • • •	
2461	ATAGTAATGAGTTTAAAAATCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTTTG	2520
		0500
2521	TACCAGTGAATTTTTGCAATAATGCAGTATGGTACATTTTTCAACTTTGAATAAAGAATA	2580
	•	
2591	сттсаасттстсаааааааа 2600	

Fig. 7/1

SUV39H Länge: 2732 bp (kodierend: 45 - 1284)

1	TCGC	GAG	GCC	GGC	TAG	GCC	CGA	ATG	TCG	TTA	GCC	GTG	GGG	Baaa	GAT	GGC	GGA	AAA	TTT	AA
								•							M	A	E	N	L	K
:1	AAGG	<u>ሮምር</u>	'C' A C'	CCT	CTC	:ጥጥር	Caa	CTC	Thirt C	•	(C) N	ጥሮጳ	درس	የረඋኦ	CCA	CCT	Стс	ጉርር	_ር ር ሞ	22
•														Q						
1	CCAA	GC T	CTC	:CTG	CCC	TGC	· CCT	CGG	TAT	CTC	TAA	GAG	GAJ	ACCT	CTA	TGA	CTT	TGA	AGT	cG
	K	L	s	С	P	A	L	G	I	s	K	R	N	L	Y	D	F	E	v	E
1	AGTA	CCT	GTG	CGA	TTA	CAA	GAA	GAT	,CCG	CGA	ACA	GGA	AT?	` \TTA	CCT	GGT	GAA	ATG	GCG	TG
	Y	L	С.	D	Y	K	K.	I	R	E .	Q	E	Y.	Y	L	v	. K	W	R	G
1	GATA																			
	Y	P	D.	S	E	S	. T	W	E	P .	R	Q	N	L	K	С	v	R	I	r
1																				
	K	Q	F	H	K	D	Г	E	R	E	L	L	R.	R	H	H	R •	S	K	T •
1	cccc																			
	P	R	H	L	D	P	s	L	A	N	Y	L	v.	Ω.	K	A	. K	Q	R	R •
1	GGGC	GCT	.cce	TCG	CTG	GGA	GCA	GGA	GCI	CAA	TGC	CAA	GCG	CAG	CCA	TCT	GGG	ACG	CAT	CA
	A	L	R	R	W	E	. Q	E	L	N •	A	K	R	S	H	L	G	R	I	T .
1	CTGT																			
	V	E	N	E	V	D	L	D	G	P	P	R	A	F	V	Y	I	N	E	¥
1	ACCG	TGT	TGG	TGA	GGG	CAT	CAC	CCI	CAA	LCCA	GGI	GGC	TGI	rggg	CTG	CGA	Gtg	CCA	GGA	CT
	R	V	G	E	G	I	. T	L	N	Q ·	V	A	v	G	С	E	с	Q	D	с
1	GTCT	GTG	GGC	ACC	CAC	TGG	AGG	CTG	CTG	CCC	GGG	GGC	GTC	CACT	GCA	CAA	GTT	TGC	CTA	CA
	L	W	A	P	T	G	G	С	С	P	G	A	S	L	H	K	F	A	Y	N
1	ATGA	CCA	rece	CCA	GG1	CCG	GCT	TCG	AGC	XGC	GC1	'GCC	CAT	CTA	CGA	GTG	CAA	CTC	CCG	CT
	Đ	Q	G	Q	V	R	L	R	A	G	L	P	I	Y	E	С	N	S	R	C
1	GCCG	CTG	CGG	CTA	\TGA	CTG	ccc	:AAA:	TC	TG1	'GG1	'ACA	GAI	AGGG	TAT	'CCG	Ata	TGA	CCT	CT
														G						
1	GCAT	C TP-T	·	Car	eci	TGA	· TGG	cc	TGG	·	333	CGT	· ·CC	Cac	CCT	יככז	GAZ	CAT	ሞርር	ca
-			R											T						
1	AGAA	CAG	CTI	CG1	CAT	rgga	GTA	CG1	rgge	Baga	\GA1	CAT	TAC	CCTC	AGA	LGGA	GGC	:AGA	GCG	GC
1	GGGG																			
	G	Q	I	Y	D	R	•	G	A	Ī	Y	L	F	D	L	D	Y	V	E	D
1	ACGI	GT	CAC	CG1	rgg <i>j</i>	\TGC	:CGC	CT?	ACT	ATGO	CAZ	CAT	CT	CCCZ	CTI	TG1	CA	LCCA	CAG	TT
	V	Y	7	V	D	A	. A	Y	Y	G •	N	I	S	. H	F	V	N	H	S	C ·
21	GTG#																			
	D	P	n	L	Q	V	Y.	N	V	F •	I	·D	N		D	E	R	L	P	R
1	GCA'I																			
	I	A	F	F	A	T	R	T	I	R	A	G	E	E	L	T	F	D	Y	N

Fig. 7/2

1141	ACATGCAAGTGGACCCCGTGGACATGGAGAGCACCCGCATGGACTCCAACTTTGGCCTGG	1200
	MQVDPVDMESTRMDSNFGLA	·
1201	CTGGGCTCCCTGGCTCCCTAAGAAGCGGGTCCGTATTGAATGCAAGTGTGGGACTGAGT	1260
	G L P G S P K K R V R I E C K C G T E S	
1261	CCTGCCGCAAATACCTCTTCTAGCCCTTAGAAGTCTGAGGCCCAGACTGACT	1320
	CRKYLF*	
1321	TGAAGCTACATGCACCTCCCCACTGCTGCCCTGTCGAGAATGACTGCCAGGGCCTC	1380
1381	GCCTGCCTCCACCTGCCCCACCTGCTCCTACCTCTACGTTCAGGGCTGTGGCCGTG	1440
1441	GTGAGGACCGACTCCAGGAGTCCCCTTTCCCTGTCCCAGCCCCATCTGTGGGTTGCACTT	1500
1501	ACAAACCCCCACCCTCAGAAATAGTTTTTCAACATCAAGACTCTCTGTCGTTGGGA	1560
1561	TTCATGGCCTATTAAGGAGGTCCAAGGGGTGAGTCCCAACCCAGCCCCAGAATATATTTG	1620
1621	TTTTTGCACCTGCTTCTGCCTGGAGATTGAGGGGTCTGCTGCAGGCCTCCTCCTGCTGC	1680
1681	CCCAAAGGTATGGGGAAGCAACCCCAGAGCAGGCAGACATCAGAGGCCAGAGTGCCTAGC	1740
1741	CCGACATGAAGCTGGTTCCCCAACCACAGAAACTTTGTACTAGTGAAAGAAA	1800
1801	TGGCCTACGGGCTGAGGCTGGTTTCTGCTCGTGCTTACAGTGCTGGGTAGTGTTGGCCCT	1860
1861	AAGAGCTGTAGGGTCTCTTCTCAGGGCTGCATATCTGAGAAGTGGATGCCCACATGCCA	1920
1921	CTGGAAGGGAAGTGGGTGTCCATGGGCCACTGAGCAGTGAGAGGGAAGGCAGTGCAGAGCT	1980
1981	GGCCAGCCCTGGAGGTAGGCTGGGACCAAGCTCTGCCTTCACAGTGCAGTGAAGGTACCT	2040
2041	AGGGCTCTTGGGAGCTCTGCGGTTGCTAGGGGGCCCTGACCTGGGGTGTCATGACCGCTGA	2100
2101	CACCACTCAGAGCTGGAACCAAGATCTAGATAGTCCGTAGATAGCACTTAGGACAAGAAT	2160
2161	GTGCATTGATGGGGTGATGAGGTGCCAGGCACTAGGTAGAGCACCTGGTCCACGTGG	2220
2221	ATTGTCTCAGGGAAGCCTTGAAAACCACGGAGGTGGATGCCAGGAAAGGGCCCATGTGC	2280
2281	AGAAGGCAAAGTACAGGCCAAGAATTGGGGGTGGGGGAGATGGCTTCCCCACTATGGGAT	2340
2341	GACGAGGCGAGAGCCCTTGCTGCCTGCCATTCCCAGACCCCAGCCCTTTGTGCTC	2400
2401	ACCCTGGTTCCACTGGTCTCAAAAGTCACCTGCCTACAAATGTACAAAAGGCGAAGGTTC	2460
2461		2520
2521	CCTGACTACCTGTGCCCAGAGTGCCCCTACATGAGACTGTATGCCCTGCTATCAGATGCC	2580
2581	AGATCTATGTGTCTGTGTGTGTCCATCCCGCCGGCCCCCAGACTAACCTCCAGGCAT	2640
2641	GGACTGAATCTGGTTCTCCTCTTGTACACCCCTCAACCCTATGCAGCCTGGAGTGGGCAT	2700
2701	CAATAAATCAACTCCACTCAAAAAAA 2732	

Fig. 8

EZH2	 1844						GTGTCCTGCA	1893
EZH1	 1				 CTGGGACT		GTTTCCTGTA	50
		1 1111111	11 11 11	111-14	11 111	H H I	GCTGGCACCA	
	51	AAAACTGC	AGCATCCA	GCGTGGAC	TTAAGAAG	CACCTGCT	GCTGGCCCCC	100
	1944						TGCAGAAAAA 	
B52	 101						TGCAGAAGAA	
BJZ	1994	· · · · · · · · · · · · · · · · · · ·	video (* 5.9988668866	b0000000000000000000000000000000000000	2000 (2000)	ee industrialis	GATGAAGCTG	ē .
	151	· * * * * * * * * * * * * * * * * * * *	50000° "AUGSSOSSOSS		**************************************	500 v 1 100000005		
	2044		de and an expension of the	and the addition of	the address of the Contract of		TCTGTTCAAC	•
	201						CCTCTTCAAC	
	2094		100000 1 10000000000000000000000000000	000000000000000000000000000000000000000	.5 40.000	deren Considerati	ACAAAATTCG 	
	251						ACAAAATTCG	
	2144		A 600 CO. 1. 4 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6	000040006000000000000000000000000000000	AAACTGCT	67 3,466,4	GTT, ATGAT	2191
	301	X	- 2000 - 1.20000000000	0.220000000000000000000000000000000000	. Marchard 2000 (2000)	00.7	GGTGAGTCCC	350
	2192				TATTTTG		GCCATCCAGA	2241
	351		* */ * *************	59.C1.V40300000000 111	- NOON CONTRACT - AND	994 : 170000 Nova	CTGTGATTTC	400
	2242			000000000000000000000000000000000000000	200000000000000000000000000000000000000	AGCCAGGC	TGATGCCCTG	2291
	401	27 70 2 200	 STTGAACA		CONTRACTOR COME. COME CONTRACTOR	TATCTTT	GTCCAAAGAT	450
	2292	AAGTATGT					* 2330	
	451	AATCATGA	TTAATATC	I I TGGTATCI	l l NTTTTAGGC	IIII CCCTCTC	489	